

ATTACHMENT A

Response to EPA: Hestmark letter 8HWM-FF received 10/25/93

1. To determine the appropriate background and operable unit populations for comparison, we understand that some matching of the two populations is done by geologists and chemists. Data for an analyte in a non-background area are grouped according to a combination of background classes which represent independent background populations. A table that cross references the operable unit populations and the background populations will be provided.

Concur. The strawman has been changed to require tables that cross-reference OU media to background media.

2. A more explicit statement of the null hypothesis that is being tested will be included. In addition, a fixed p value of 0.05 will be used for each of the inferential statistical tests as written in the strawman proposal. There was some inconsistency in what was written in the proposal and what was stated in the meeting regarding the p value. A fixed value of 0.05 is what we will accept.

Concur. The strawman states that p values must be less than or equal to 0.05 to demonstrate a significant difference from background. Footnote 3 on page 5 of the strawman, which was not clear on this point, has been deleted.

3. All references to comparison of background and operable unit populations for organics will be removed. Background comparisons apply to inorganics and radionuclides only.

Do not concur. Although background comparisons for organics are not commonly used, there are instances when it may be applicable, in which wide-ranging organic contamination is due to non-site-specific anthropogenic sources. We want to retain the option of performing background comparisons for these organics, when geochemists or geologists determine that it is applicable to do so. In these instances, we will retain the burden of proof, and the applicability of the comparison will be subject to EPA and CDH approval.

The strawman has been rewritten to state that background comparisons for organics will be done on a limited, case-by-case basis, subject to EPA and CDH approval.

4. The use of professional judgement in interpreting the results of the graphical displays and statistical analyses will be limited to consideration of spatial distribution, temporal distribution, and pattern recognition concepts. The strawman proposal included five additional criteria. These will be deleted in the final implementation document.

Concur. The five criteria (intermedia interactions and geochemical processes, not an expected contaminant, blank data, regional background range, and influence of field activities) have been deleted.

5. The non-background population is defined as the entire operable unit remedial investigation set. The data aggregation for the purpose of background comparison will be done within the area defined by the operable unit boundaries.

Concur. Analysis will be done on an OU-wide basis.

6. The attached flowchart, "Background Comparison Methodology", distributed at the meeting will be clarified. It is EPA's understanding that all the data sets will undergo the hot measurement test and the battery of inferential statistical tests (Gehan, Quantile, Slippage, and T-Test) provided the data satisfies the conditions stated in the strawman and on the flowchart. If any one of these tests, including the hot measurement test, shows significance, the analyte will be further considered, using professional judgement, as a contaminant of concern. The flowchart would benefit from the addition of decision blocks after each test indicating the next step if significance is demonstrated or not.

Clarification. The chart "Background Comparison Methodology" attached to EPA's memo is not the same as that distributed at the September 29, 1993 meeting and contained within the strawman proposal. The difference is that nonparametric ANOVA tests are given as options to the Gehan test in the chart within the strawman proposal. Because the Gehan method is not standard and will therefore incur practical liabilities (e.g., the method has not been adequately tested and verified, preliminary usage shows it to require excessive man-hours, and subcontractors will need to be instructed in its use), we want to retain the option of performing standard nonparametric ANOVA testing, using the Wilcoxon or Kruskal-Wallis tests, instead of the Gehan test.

Additional clarification. All tests will be performed, if applicable, regardless of whether other tests demonstrate significance.

Concur with the need to redo the flowchart. This has been done.

6. (continued) We also have some specific questions that need to be addressed in the final document:

a. What happens to data which is carried through the slippage test but does not qualify for the t-test?

Clarification. The data that do not qualify for the t-test will be routed to the "At Least One Test Significant?" block. The flowchart has been revised to show this.

b. What is the basis for the 20% detect value as the criteria for the Quantile test? How does this criteria relate to the criteria for applying this test as stated in Dr. Gilbert's report on page 20?

Clarification. Dr. Gilbert's method proposed looking up tabulated values for n and r parameters. The quantile test could be correctly applied only if the largest n values

were all detects. Our statisticians have stated that, typically, this restriction equates to the largest 20% or less of the combined sample sizes being detects, and recommend using a flat 20% to simplify application.

- c. What is the basis for the criteria of $N > 20$ value for background and operable unit data?

Clarification. Our statisticians derived this value from application of the Central Limit Theorem for a two sample problem. If both samples have $N=20$, then there will be 38 total degrees of freedom, which will permit assumptions about the distribution.

7. EG&G's claim that these impacts [of implementing Dr. Gilbert's recommendations] could range from \$30,000 up to \$120,000 per operable unit is not supported by the information provided. In fact, it appears that there is some evidence that implementation will not negatively impact costs or schedules.

Do not concur. EG&G had provided reasoning behind these estimates in memo 93-RF-11078 (STATISTICAL METHODOLOGY FOR BACKGROUND AND COMPARISONS AT THE ROCKY FLATS PLANT - NMH-463-93) dated September 15, 1993. Because the Gilbert method requires additional work, there will be cost and/or schedule impacts.

In addition to the impacts mentioned above, significant cost impacts are anticipated to result if the Gehan method is used. For OU11, approximately 200 hours were required to perform the Gehan test, when less than 40 hours would have been sufficient to perform standard ANOVA testing.